

2009 年度修士論文要旨

## 系列パターンマイニングのためのカスケードモデル

関西学院大学大学院理工学研究科

情報科学専攻 岡田研究室 吉 川 芳 浩

顧客の購買行動，Web ページの閲覧履歴，医療データなどのアイテムの発生順序に意味を持つ系列データがある．これらの系列データから，頻出する系列パターンを見つけ出す手法を系列パターンマイニングと言い，これまでにアルゴリズムがいくつも紹介されている．データによっては頻出のアイテムセットを発見する手法に比べ大変有用な情報を得られることが期待されているが，予想できるような系列パターンしか得られない場合もある．そこで本研究では，クラス属性の識別力が高い特徴的な系列パターンを発見する新しい系列パターンマイニング手法を提案する．

本研究室ではクラス相関ルールの拡張であるカスケードモデルを提案している．これは特徴的ルール導出と呼ばれ，現在データマイニングの主要な技法の1つとなっている．そこで，これまで扱えなかった系列データを説明要因としてルール導出を行うためにカスケードモデルを拡張した．また頻出の系列パターン導出には系列パターンマイニングの代表アルゴリズムである **PrefixSpan** を用い，得られたパターン間で急激なクラスラベルの分布変化を検出して，ルールとして表現するシステムを実装した．このシステムを利用しオンラインショップの Web 閲覧履歴を対象データとし，購入者と非購入者を識別できるような特徴的パターンの発見を目的とする．